Ocean Science Discussions

Ocean Sci. Discuss., referee comment RC1
https://doi.org/10.5194/os-2022-11-RC1, 2022

Comment on os-2022-11

Anonymous Referee #1

Referee comment on "Ensemble quantification of short-term predictability of the ocean dynamics at a kilometric-scale resolution: a Western Mediterranean test case" by Stephanie Leroux et al., Ocean Sci. Discuss., https://doi.org/10.5194/os-2022-11-RC1, 2022

Review of os-2022-11

General comments:

This interesting manuscript focuses on the predictability of small scales in realistic ocean models kept "on track" by data assimilation (although the manuscript does not contain assimilation results). In particular, it proposes a rather novel methodological approach to relate forecast uncertainties to initial uncertainties in the fields, and presents some results quite convincingly in the context of a particular experimental protocol based on a set of 2D "displacements". The topic is scientifically relevant and important and the scientific quality is good, but the focus, clarity and precision could sometimes be greatly improved. I have no reservations about the statistical/probabilistic methodologies implemented, and the results are valid and interesting, but I am not convinced of their generality given the particular experimental protocol (type of uncertainties considered, seemingly "fixed" scale, number of members, etc.): the limits of the ensemble generation approach, and thus the scope and validity domain of the results, should become more apparent. This manuscript should eventually be accepted for publication, but perhaps not quite in its present form.

Specific comments:

The style of the introductory and methodological sections is sometimes rather "literary" and "rhetorical", convoluted to the point of being imprecise (an example: see the comment "lines 56-68" below) -- the approach is often introduced by invoking much more general and theoretical concepts than necessary. On other occasions, the text does not contain enough information or loses the reader. I would recommend (1) adopting a much more "direct", "factual", "scientific" style throughout the text, and (2) improving precision and conciseness. For example, when describing a methodology, the description of what

was done in practice could be presented first, accurately and completely (and not in three different places, such as the perturbation scheme in sub-sections 2.2 and 3.1 and Appendix A); then the validity and scope of the approach, including the wider context, can be discussed, not the other way around.

However, as the ms. progresses, the style improves, especially in the description of the results, which is often adequate.

The definition of predictability scores (in particular CRPS and predictability diagrams), and the way in which statistical calculations are carried out using all members of the ensemble in turn as a reference (reminiscent of generalised cross-validation) are two aspects of the work that could be generalised to problems beyond the particular experimental protocol. I was particularly interested in the dispersion of the CRPS estimates across the 20 cases (Figures 7,8) -- I would be curious to know what they look like with only the reliability CRPS component or only the resolution component (the latter possibly giving access to a form of feature-based predictability, i.e. based on whether a particular forecast eddy is present across the members). The decorrelation score is interesting and also seems to be quite general. The location score is of course more related to the particular type of uncertainties in the study.

While section 4 is solid, one should keep in mind that the predictability analyses are conducted within a very specific experimental protocol: that of pseudo-random perturbations based on 2D "displacement" at small scales (10 grid points), having a direct impact on horizontal advection and pressure gradient at these scales (and indirectly on other dynamical processes). (I know that "displacement" is probably not the right term as you are perturbing the metrics of the model operator, not the grid, but you could give this word that definition in your ms). That's OK, but in retrospect I probably would have liked a more honest introduction and summary framing the study more clearly in the particular experimental protocol (e.g. as described in subsection 2.2 from line 118). Indeed, the results in Section 4 could be very different for other forms of uncertainty. The conclusion is not careful enough in this respect: its first sentence ("The overall aim of this study...") promises too much in relation to the very real and effective work that has been done.

In addition, a limitation of this work that is not mentioned in the conclusion is that the correlation scale of the displacements is set (if I understood correctly) at 10 grid points. So, if I understand correctly, this is a predictability analysis study for a 10 grid point noise. Would a smaller or larger scale noise behave in the same way? What about pseudo-random correlation scales? However, I'm not sure I understood correctly, since the conclusion quotes "10 km wavelength" and not "a scale of 10 grid points" -- which is quite confusing. Similarly, the tenfold use of a Laplacian filter is mentioned -- even more confusion.

Twenty members is a small size for an ensemble, again a topic not addressed by the conclusion. It is not clear whether we should interpret the discussion in 3.2.3 as evidence that 20 members are "sufficient" for the subsequent predictability study? What about the representation of spatial covariances with 20 members? (These generally converge more

slowly than the variances). Also, what is the impact of the ensemble mean, and is it taken into account?

This is a scientific paper. Therefore, the emphasis on CMEMS, which is cited several times, and which also comes as the "last word" in the conclusion, seems out of place. Such a study is of interest to all ocean forecasting systems. If appropriate, CMEMS can be mentioned in the acknowledgements.

Individual comments:

lines 56-58: This appears as a purely rhetorical statement, but perhaps I did not understand what was meant. Models and assimilated observations have errors which do impact the forecasts, we know that. Also, how can model instabilities be used to produce a valuable forecast?

lines 62-63: "initial uncertainties because observation resources are limited": yes, but observations have errors too; and in an assimilating system initial errors are also due to the whole history of all types of errors up to then.

The introduction has no references on probabilistic skill scores.

line 98: "initiated" -> "initialised"

section 2.1: Which scales can be accurately modelled by MEDWEST60? It is important to have those in mind in relation with the perturbation scales which you will use. Also, in the Mediterranean the internal Rossby radii are quite small.

lines 109-110: "In this context..." -> "In a purely deterministic approach..." to improve clarity. But still, you are missing modelling errors here (parameterisation, numerical schemes, missing physics).

lines 148, 151, in Table 2, etc: "probabilistic model" -> "stochastic model"

line 154, legend of Figure 2, etc: "grid size", "size of the model grid" -> "grid spacing" or "mesh spacing". Also what is the distribution law used for the perturbations? (If a non-compact support law is used, did you use an upper bound for the displacement?)

lines 163-164: "It does rely...": I do not understand the sentence (you wrote the opposite two sentences before). Also: part of this paragraph is descriptive, and part is a discussion in anticipation for another discussion in chapter 4: it is not good to mix everything because you'll get the reader lost.

Table 2: I do not understand what "identical" initial conditions mean. I would have thought that the spun-up fields would be pseudorandomly displaced using the 20 samples of the displacement fields (for each of 1%, 5% stdev), hence yielding 20 *different* initial conditions across the ensembles.

lines 182-183: I am a bit confused. The "displacement" is variable, with stdev = 1%-5% of the mesh spacing, but the displacement correlation scale is fixed to exactly 10 gridpoints. Therefore I do not understand the words "on the order of".

Figure 3: It might be interesting to have a zoomed version on the right (perhaps just for low wavenumbers) to be able to see something.

I did not have time for a full second reading and hence for further individual comments, sorry.